

Performance measurement and evaluation in arts management

A meta-synthesis

LETICIA LABARONNE*

Center of Arts Management, Zurich University of Applied Sciences; PhD Student, WÜRTH
Chair of Cultural Production, Zeppelin University

Abstract

This study analyzes the performance measurement and evaluation literature in (non-profit) arts management by conducting a meta-synthesis, which is both a process and a product of explorative scientific enquiry. Meta-syntheses go beyond the well-known procedure of literature reviews, often used to summarize the current state of knowledge in a particular field, in that they produce formal integrations that offer novel understandings of the reviewed literature. This article presents, in the results section, the consolidated understandings under four thought-provoking titles: First, 'Third time's a charm?' describes the development phases identified in the body of performance measurement and evaluation literature. Second, 'A paradigm on the move?' refers to the dominance of the positivist research tradition and signals the quest for alternative approaches. Third, 'Pride and Prejudice' illustrates that the international literature tends to emphasize the benefits and learnings of performance measurement and evaluation practices (pride) while a rather prejudicial attitude is observed among the German-written literature. Fourth, 'Good Cop, Bad Cop' takes the debate to a higher level by thematizing the interplay between arts management research and cultural policy. The conclusion discusses the four titles through the lens of system theoretical discourse.

Key words

evaluation; method development; arts administration

1. Introduction

The nonprofit sector is increasingly permeated by expectations of measurement (PATON 2003) and today nonprofits are held accountable for what they deliver (EBRAHIM 2010). Demands for accountability may be broken down into four broad, far from comprehensive, categories: accountability for finances, governance, performance, and mission (BEHN 2001). These demands have arisen mostly from external stakeholders such as donors, public funders, business sponsors, and the general public (CAIRS et al. 2004). However, nonprofits themselves have also taken a proactive step, anticipating future demands, as they struggle to im-

* Email: lael@zhaw.ch

prove services, compete for resources and preserve credibility and legitimacy (WILLIAM/BODWIN 2007). Nonprofit theory and practice have responded to this 'performance measurement imperative' with different approaches (GSTRAUNTHALER/PIBER 2012). Yet, as the lack of a profit orientation leaves nonprofits without a common goal, an overarching performance measurement or evaluation system is missing (Ibid) and the puzzle of how to measure and evaluate performance remains a recurrent issue in nonprofit literature and practice (CAIRS et al 2004).

1.1 The emergence of evidence-based cultural policy

Every development no doubt happens embedded in a specific historical context. This holds true for the emergence of the issues of performance measurement and evaluation in the arts and cultural sector. With the advent of Thatcherism and Neoliberalism in the late 1970s, a strong political commitment to markets and competitive individualism (HUTTON 1995) led to a re-examination of public sector programs, including the arts and cultural sector in the UK (GILHESPY 1999). Furthermore, in the United States (US) the emphasis on performance measurement and evaluation was motivated by a significant decrease in public funding for the arts and culture in the mid-1990s (BROOKS 2000). Today, research into the value and impact of the arts constitutes a core function of the National Endowment for the Arts (2014), the federal arts funding agency in the US. Likewise, the current research activities of the Arts Council England (2014), the National Development Agency, focuses on monitoring the performance and impact of art and cultural activities as well as providing evidence for policy-making. In German-speaking countries, in which the arts sector – in contrast to the US¹ – is characterized by high levels of direct government support, cultural policy is also increasingly evidence-based. In Switzerland, for example, the New Culture Promotion Act (2011) stipulates that the effectiveness of cultural policies must be revised regularly.

1 In the United States only about 13 % of direct arts support comes from public funding and only 9 % from the federal government. The rest comes from earned revenue and private sources of funding. That is, the large proportion of arts funding is 'indirect' in terms of tax deductibility of gifts for NPOs. For every dollar of direct support, the USA provides about 14 dollars of indirect support (THE NATIONAL ENDOWMENT FOR THE ARTS 2014). Hence, ACOs in the US play a more important role on individual contributions and fundraising (WEINSTEIN/BUKOVINSKY 2009: 45).

Against this background, arts and cultural organizations (ACOs)² are faced with a stronger political logic demanding them to focus on tangible results and to prove how they add value to their communities (GSTRAUNTHALER/PIBER 2012). These demands are less directed at the positive externalities of the arts than they are at the associated opportunity costs of public money; that is, whether the benefit is greater than the benefit that could be gained from alternative spending scenarios (COWEN 2008). The latter has been encouraged by the discourse in many Anglo-Saxon countries about the apparent power of the arts to counter social exclusion or enhance community life (JOWELL 2004). While some lament this “instrumentalization of the arts” (BELFIORE/BENNETT 2010: 122), others have commented on the idea of policy attachment, whereby the arts, as a less influential policy area, have attached themselves to economic and social agendas to get a stronger political support (GRAY 2002). Moreover, among private funding bodies, evaluations are also increasingly regarded as a valuable tool to assess the achievement of objectives (SCHÖBER et al 2012). While this instrumental rationality is widespread in the performance measurement and evaluation discourse in the arts and cultural sector of much of the West, it is by no means uncontested. Some researchers argue that resources would be better spent trying to understand the arts themselves (HADIDA 2015).

1.2 Terminology delimitation

A short literature review of the body of knowledge in arts management indicates that the terms ‘performance measurement’, ‘performance evaluation’ and ‘evaluation’ are being used in a rather undifferentiated manner. Allègre Hadida (2015), who conducted a critical literature review about performance in the creative industries, argues that the performance definitions and measurement approaches of practitioners, experts and academic researchers might differ, but assumes that these differences are marginal. One might agree with Hadida, particularly given that performance measurement and evaluation are related and often

- 2 In his effort to define arts and cultural organizations (ACOs), KAPLE (quote in WEINSTEIN/BUKOVINSKY 2009) concludes that there is no universally accepted definition, among other things since many organizations operate informally and under an umbrella institution. The present study draws, however, on a rather narrower definition of ACOs, referring to those organizations operating on the not-for-profit corner of the arts and culture sector. In limiting the study to a narrower segment The author aims to provide a sharper analytical focus.

complementary. Nonetheless, there is a difference between these concepts (NIELSEN/EJLER 2008).

According to Schober et al. (2012), performance measurement is a continuous process in which targets and goals are defined and then operationalized in measurable indicators. It serves as a tool for exercising control and for ensuring accountability, focusing primarily on efficiency and effectivity rather than on measures of impact. Evaluations, on the other hand, tend to focus on the content of the activities of an organization or project and thus can produce findings that relate to impact. In order to do that, evaluations often go through complex processes that need a long-term horizon. Hence they are not useful for controlling daily operations.

There are three basic types of evaluations (GETZ 2005): (i) formative evaluations (ex-ante) are often used for program conception or pre-event assessment, (ii) process evaluations (on-going) assess the process and thus they are linked at the organizational level with performance measure and quality management systems, and (iii) outcome or summative evaluations (ex-post) measure of outcome and impact. Today the evaluation literature in the nonprofit sector acknowledges the difficulty in measuring the intangible, so the current discourse about impact measurement instead uses the term impact analyses often (SCHOBBER et al. 2012).

1.3 The challenges of assessing artistic and cultural activities

As a specific type of nonprofit, ACOs are also involved in the performance measurement and evaluation debate – particularly in terms of defining performance (HADIDA 2015) as well as developing assessment approaches sensitive to their uniqueness and complexity (SOREN 2000). The challenges in the (nonprofit) arts and cultural sector can be partly explained by the perspective of art as a system put forward by theorists such as Arthur Danto (1964), Niklas Luhmann (1997), Howard Becker (1982), and George Dickie (1984). From this perspective, art is not only a social construction but also a social product. Defining performance in the arts and cultural sector is thus all the more difficult, as the product (output resp. outcome) often exhibits characteristics of merit, public or semi-public goods, and might involve intangible individual or collective experiences that are not captured through private transactions (HADIDA 2015).

Furthermore, the evaluation of art might be influenced by specific conventions strongly related to its context, and thus the context-relat-

ed uniqueness of artistic accomplishments should be taken into account (BECKER 1982). In addition, as ACOs are embedded in a network of various stakeholders with different interests (BROOKS et al. 2002), the performance measurement and evaluation of artistic and cultural activities involves a balancing act between aesthetic purposes and market orientation (LAMPEL et al. 2000). Not only are outputs aimed at different stakeholders simultaneously, but the results also may span over several years, as it is the case in heritage conservation activities (FINOCCHIAIRO/RIZZO 2009).

1.4 Problem Statement and Study Aim

The issues of performance measurement and evaluation in the (non-profit) art and cultural sector are recurrent topics in the arts management literature. Therefor numerous approaches have been proposed.³ Yet, research has barely analyzed the body of knowledge against the background of the different artistic contexts as well as the broader environments in which the performance measurement and evaluation of ACOs takes place. This is all the more surprising not only considering the unique contextuality of artistic and cultural activities but also taking into account the different systems (e.g., artistic, economic, political) and its complex networks of actors, institutions, values and interests in which ACOs are embedded. Funding bodies are key stakeholders in these networks. As such, they exert an influential role in the framing of the analysis that underpins the construction of evaluation practices (CLEMENTS 2007). Giving the different funding structures for the arts and culture across regions (e.g., US vs. Continental Europe resp. German speaking-area), the question of how funding bodies shape the evaluation process seems particularly relevant in cross-regional comparison. The latter also because the act of measuring itself, as a social construct, is never value-neutral but always embedded in a particular context (WIMMER 2004). Against this background this study aims to analyze the body of the performance measurement and evaluation literature across and against artistic disciplines, institutional settings, and regions. The overarching research question aims at learning from different studies in order to identify patterns, emerging trends, or any other interesting relationships.

3 See e. g. GILHESPY (1999, 2001); KRUG/WINBERG (2004); WEINSTEIN/BUKOSVINSKY (2009); GALLOWAY (2009); ZORLONI (2012); BADIA/DONATO (2013); BETZLER (2015).

2. Methodology

The intention of the study is not to aggregate data on performance measurement and evaluation projects in order to compare study results. The focus of the analysis is rather on the different approaches proposed in the academic literature, such as instruments, methods, frameworks or theoretical models, and on the identification of patterns of similarities and differences or any other noteworthy relationship that may come to light. In order to do this a meta-synthesis is conducted. Whereas literature reviews engage critically and/or discursively with earlier work to summarize the current state of knowledge in a particular field, meta-syntheses go through and beyond both narrative and systematic reviews to produce formal assimilations or integrations of the reviewed works that offer novel understandings or insights (SANDELOWSIKI/BARROSO 2007).⁴

A meta-synthesis, as both a process and a product, is a scientific inquiry that can be understood as a qualitative counterpart to the quantitative meta-analysis (SANDELOWSIKI/BARROSO 2007). Because the findings are reconceptualized (DOYLE 2003), the validity of a meta-synthesis does not depend upon the replication but rather upon the critical interpretation. Furthermore, meta-studies can also offer a historical staging and/or explanatory context and assist researchers in exploring differences as well as similarities across settings, sample populations, disciplines and methodological approaches (HANNES/LOCKWOOD 2012).

There is an increasing variety of methodological approaches that is used to conduct a meta-synthesis, including meta-ethnography, qualitative meta-synthesis, realist review, thematic analysis, critical interpretive synthesis, and framework analysis/synthesis (LEE et al. 2015). Some methods are more suitable for producing recommendations relevant to practice and policy formulation while others are more suited to theory and model development. Meta-ethnography is perhaps the most well developed and established method for conducting a meta-synthesis and one that clearly has its origins in the interpretive paradigm (POPE/MAYS 2006). Meta-ethnography was introduced by Noblit and Hare (1998), who devised a seven-step iterative process for conducting research inquiry: getting started, deciding what is relevant to the initial

4 For a systematic comparison between meta-analysis, traditional literature review, and meta-ethnography, see DOYLE (2003:324).

interest, reading the studies, determining how the studies are related, translating the studies into one another, synthesizing translations, and expressing the synthesis.

The formulation of the question is framed by the aforementioned problem statement. However, in this type of explorative project, it is not a simple sequential procedure but rather an iterative one that is likely to be sharpened as deeper research questions emerge over the course of the study. Thus a meta-synthesis cannot be conducted as a linear process. It involves experimenting with the method, as phases may overlap, even run in parallel, and the work undertaken may not always advance towards the final synthesis (LEE et al. 2015).

In light of the above, this study was approached in a way that allowed the research design to gradually emerge as the inquiry evolved, while important methodological decisions, such as determining the point of saturation or selecting issues for further scrutiny, were taken as the study evolved (DENZIN/LINCOLN 2011: 671). In order to preserve methodological transparency, careful documentation was carried out (audit trail) and efforts were undertaken to make the underlying assumptions clear and the decision-making procedures transparent.

The research design is framed alongside Noblit's and Hare's seven-steps, as they remain the primary organizing device for conducting a meta-synthetic inquiry, but was adapted and operationalized in a way that serves the overarching research question.⁵ The study started with an inductive approach in order to identify relevant questions and then moved on to a more deductive approach aimed at generalizing the exploratory findings. In order to do that, it used a classic mixed-methods design, which combines probability and purpose sampling in order to increase validity and credibility (PATTON 2015: 65). By doing things in this way, the sample 'aligns' (PATTON 2015: 265) with the emergent research questions in a continuing negotiation process involving trade-offs between breadth and depth. Unlike (quantitative) meta-analyses, samples in meta-ethnographies are purposive rather than exhaustive, because the purpose is interpretive explanation and reconceptualization rather than prediction (DOYLE 2003).

5 For a review of how a range of different authors describe doing meta-ethnography, see LEE et al (2015: 338).

2.1 Search strategy

The study considers literature published in English in international journals as well as works written in German⁶ that investigate the topics of performance measurement and evaluation in the non-commercial corner of the arts and cultural sector. A systematic bibliographic search using the databases JSTOR and EBSCO was conducted. In addition, the Karlsruher Virtueller Katalog (KVK), the Network of Libraries and Information Centers in Switzerland (NEBIS) and the Austrian Network and Service of Libraries (OBVSG) were used to access the body of literature in German-speaking countries. In these areas, doctoral dissertations, and sometimes masters theses, are often published as monographs and these are regarded as a relevant form for disseminating knowledge. Given that a failure to conduct a sufficiently exhaustive search can jeopardize validity (SANDELOWSKI/BARROSO 2007), the search strategy encompassed not only peer-reviewed articles but also significant monographs and books in an effort to include fugitive or grey literature.⁷

The choice of terminology for the literature search is self-explanatory. The core keywords ‘performance’, ‘performance measurement’, ‘performance evaluation’ and ‘evaluation’ were firstly used as umbrella terms under which the different approaches to these topics were fit. As aforementioned, performance measurement and evaluation are different concepts, even though they are related and often complementary. This study deliberately encompasses both concepts. It did this, first, because the terms seem to often be used interchangeably, and the focus on ‘only’ one or the other concept would be a too narrow search strategy for an explorative study. Second, it did this because evaluative practices have been integrated into manifold conceptualizations of New Public Management (as well as entered the realm of the private sector) and evaluation are spreading into the key fields of managerial action with concepts such as Total Quality Management, Benchmarking, EFQM, Auditing, and Controlling (CHIARAVALLOTI/PIBER 2011). The latter underlines the premise that the concepts are (increasingly) related.

Furthermore, the search strategy combined the core keywords with the broad-based terms ‘art organizations’ and ‘cultural organizations’, which were then translated into German. The search strategy was rede-

6 The scope of German-speaking countries should be attributed to the fact that the author is a researcher active in this area; it should not suggest that comparisons across the body of literature in other regions might be less relevant.

7 This refers to potentially relevant literature that is not accessible via electronic databases (SANDELOWSKI/BARROSO 2007).

defined iteratively by incorporating key terms – such as ‘indicators’, ‘impact’ and ‘accountability’ –, which were identified from relevant papers. In addition, as the number of academic journals in arts management is growing (LINDQVIST 2012: 11) and the research area of performance measurement and evaluation in the arts and cultural sector is multidisciplinary, the research field may be broad and the body of literature may be disseminated over various journals and disciplines.

3. Results

3.1 Screening and exclusion criteria

The first screening was carried out according to the following working definition: Works (peer-reviewed articles, monographs, and book chapters), whose primary focus is the introduction or discussion of approaches to performance measurement and evaluation in the non-commercial edge of the arts and cultural sector.

The systematic bibliographic search yielded a total of 1,268 records (including duplicates), regardless of publication date (see Fig. 1). The initial screening focused solely on the titles and keywords and mostly aimed at excluding works for topical reasons (such as studies on evaluating organizational culture).

After the initial screening, the abstracts of 89 records were retrieved. In addition, given the broad range of terms in use, which sometimes are not adopted as keywords, the electronic literature search was supplemented by a more informal approach that included reference chaining and validation with experts in the field. Altogether, a further 13 abstracts were identified through informal search methods. In total, 102 records were retrieved, scanned by reading the abstracts, and examined for inclusion and exclusion criteria. The definition of exclusion criteria, in particular, was formulated in an iterative manner, as most of the decisions for inclusion/exclusion emerged during the screening process.

Following the aforementioned working definition, studies with a methodological focus – for example works that propose a new approach or discuss existing methods – were included in the sample. These had to be related to not-for-profit art and cultural activities either at an institutional, program or project level, for which some sort of public and/or private funding was obtained. Excluded from the sample were works relating to commercially oriented art and cultural activities (e.g., art-trade, commercial films, record labels, musicals). These activities are subjected

to an overarching profit goal and thus can be more straightforwardly defined and measured with traditional commercial performance indicators and assessed in economic terms. In addition, they are well researched across units of analysis and social science disciplines (HADIDA 2015).

The sample also excluded works focusing on art school education, science centers, arts or cultural activities aimed at social integration (e.g., child and family welfare), and cultural tourism, because the focus on performance measurement and evaluation activities tends to revolve around pedagogical, welfare or service concerns rather than around artistic merit and the experiences of art production and reception.

In addition, also excluded from the final sample were reports, handbooks, guidelines, textbooks (as in teaching books), and master theses, given that these works are not subjected to the same scientific requirements as peer-reviewed articles and dissertations.

From the 102 abstracts scanned, 57 works did not conform to the inclusion/exclusion criteria (see Fig. 1). In total 45 full texts were retrieved for in-depth analyses,⁸ which constituted the final sample (See Appendix).

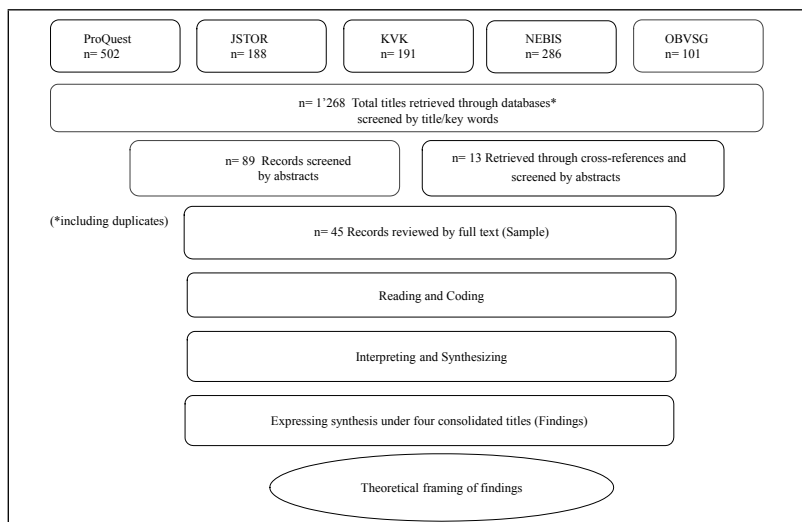


Fig. 1: *Research design*.

- 8 Even though the study engages thoroughly with each piece of selected work, the synthesis is based on the assumption that the selected works are of acceptable quality, as this paper does not address the question of critically appraising research. Critical appraisal is the “process of systematically examining research evidence to assess its validity, results and relevance before using it to inform decision” (HILL/SPITTLEHOUSE 2001: 1).

3.2 Reading and Coding

While the identification of relevant studies is an important element of meta-synthesis, the process of actively reading the chosen works, identifying the main coding themes, and extracting data is arguably the more difficult, even significant, step (LEE et al. 2015: 337). Surely it is the most time consuming. Actively reading means that reading might need to be undertaken at different points in the process, for different purposes: to appraise, familiarize, identify, extract, record, organize, compare, relate, map, stimulate, and verify, for which several techniques including annotating, coding, listing, and tabulating were used (HANNES/MACAITIS 2012).

The selected studies were read in ‘waves’, in a way that reflects the grounded-theory approach. First, each work was coded noting the approach and the context in which it has been applied or proposed in order to record contextual information to be used in the interpretations of each study. The formulation of these first codes was derived by the content of the overarching research question. Second, the works were re-read and organized into (sub-)groups, identifying common and recurring concepts as well as categories and new coding themes. At the end of the second reading wave, the following coding themes were defined and used to analyze the selected studies:

- #1 country/region/language
- #2 artistic discipline
- #3 institutional setting (e.g., financial structure, size, own production)
- #4 conceptual approach
- #5 research tradition
- #6 units of analyses
- #7 starting point/project owner/research motivation
- #8 whom are the results meant to inform
- #9 discussion about implementation
- #10 research limitations

A mix methodology involving both deductive and inductive approaches was used for the formulation of coding themes. While codes #1, #2, #3, #4, #5 were defined a priori, the rest of the coding themes emerged as the reading process evolved.

3.3 Characteristics of the selected works

The reviewed literature stretches from 1986 to 2015. As expected, when defining the search strategy, the literature sample was spread throughout a vast number of journals and, in the case of the German-written literature, publishers. However, around 20 % of the studies were published at the *International Journal of Arts Management*, not surprisingly in the track 'Measurement of Cultural Organization Performance'. The majority of the works focus primarily on instrumental and methodological issues and revolve around new or adapted approaches to performance measurement and evaluation (coding theme #4), in almost 50 % of the cases addressing ACOs in general (coding theme #2). In addition, specific approaches have mostly been developed for museums (n=11) or performing arts organizations (n=8). Further, only a few studies critically reflect on the body of literature and address a range of the (often neglected) theoretical, epistemological, aesthetic, political, and ethical challenges embedded in the evaluation process. An emerging trend can as of 2007 be seen (BELFIORE/BENNETT 2007; GALLOWAY 2009; CHIARAVALLOTI/PIBER 2011; CHIARAVALLOTI 2014; HADIDA 2015).

Many of the limitations discussed in the studies (coding theme #10) revolve around the operationalization of the artistic dimension (construct validity), regardless of the kind of approach taken. The challenges seem to concern, at the outcome level, how to capture the quality of artistry and artistic achievement and, at the impact level, how to conceptualize and measure the value creation of these activities. Whereas measuring at the level of process and output seems more straightforward, assessing the effects of artistic and cultural activities on society seems more challenging – that is, either the direct impact on the recipients (e.g., theater goers or museum visitors) or the indirect impact on the broader community (i.e., externalities).

Furthermore, the limitations also encompass issues of relevance and applicability. The body of literature on performance measurement and evaluation in the arts and culture tends to be normatively driven. That is, very few of the reviewed studies investigated how performance measurement and evaluation operates in arts management practice, for whom and by whom it is done, and how it is perceived by the different stakeholders (GILHESPY 1999; COHEN/PATE 2000; ZORLONI 2012; GSTRAUNTHALER/PIBER 2012). This observation is to a large extent coherent with the literature. The missing link between theory and practice was mentioned by Turbide and Hoskin already in 1999 and cor-

roborated by more recent works (DONATO 2008; TURBIDE/LAURIN 2009).

3.4 Interpreting and synthesizing

After coding and extracting data, the next step resides in its interpretation and synthesis. In the context of meta-ethnographies, the concept of synthesis should not be understood as an additive process, but rather as a reconceptualization across studies (DOYLE 2003). The following subsection presents the 'reconceptualizations' that emerged by looking into differences and similarities among the selected works along and across the coding themes. These understandings are consolidated through four thought-provoking titles in an attempt to highlight the findings for further discussion. Please note that not all coding themes are encompassed in the following condensed understanding as some of them, such as #1, #2 and #3 were primarily intended at providing contextual information about the sample.

Third time's a charm? Across the academic literature investigating performance measurement and evaluation in the arts and culture sector, three phases or development waves can be identified (coding theme #4).

In the beginning, performance measurement in the arts sector was limited to financial indicators such as attendance levels, number of performances or exhibitions, and earned income. This is not surprising considering that research on performance evaluation is rooted in the field of management accounting (CHIARAVALLOTI 2014). This first phase was thus characterized by quantitative indicators that supply no relevant information about artistic achievement.

In order to better convey the complexity of ACOs, a second wave of research has proposed numerous multidimensional frameworks that aim to merge quantitative indicators and quality concerns. For example, Gilhespy (1999, 2001) proposed a framework that includes a measurement of artistic excellence and innovation. Krug and Winberg (2004) proposed three dimensions for assessing effectiveness in ACOs: contribution to mission, contribution to money, and contribution to merit. More recent works have proposed the Balanced Scorecard as a tool to integrate artistic and intellectual drivers in the evaluation of ACOs, such as theaters (WEISTEIN/BUKOSVINSKY 2009) or art museums (ZORLONI 2012). In German speaking countries, a growing amount of academic literature focuses on the application of Quality Management Systems to

ACOs (TSCHÄPE/BRÜGGERHOFF 2004; SCHEYTT/ZIMMERMANN 2006; CORDES et al. 2008; ZULAUF 2012).

By the time the second development wave towards a multidimensional framework began, it was widely agreed in the literature that a comprehensive approach to performance measurement and evaluation was needed in order to represent the multifaceted nature of the phenomena in the arts and cultural sector. However, in recent years, a challenge to the idea of multidimensional frameworks can be observed, indicating the emergence of an (overlapping) third phase. This last developing phase questions the very same underlying assumptions on which the multidimensional models are built. Namely, on the premise that the fulfillment of organizational goals and the aesthetic dimension of performance can be represented with a couple of key figures. On one side, as rich qualitative data cannot be easily transformed into one-dimensional figures, there is the question of whether the aesthetic dimension can be captured by metrics (GSTRAUNTHALER/PIBER 2012). Intangible individual and collective experiences may thereby vanish as the inner value of arts and culture, such as the value reflected in the feelings of the beholder of an artistic work and those involved in its creation, remains hidden behind the representation of 'success' or 'effectiveness' (HOLDEN 2004). On the other hand, there is the issue of aggregating several components and how to merge the different rationalities that underline different financial, artistic or political logics (GSTRAUNTHALER/PIBER 2012).

A paradigm on the move? The dominance of a positivist research tradition in the discourse about performance measurement and evaluation in arts management can be observed (coding theme #5), which is coherent with the literature. Following Chiaravalloti and Piber (2011), the three categories considered by Wicks and Freeman (1998) – the way of looking at reality, the methods and techniques used, and the relevance given to ethics – were used to establish the main differences between the positivist and anti-positivist approaches.

Positivist approaches assume the context-neutrality of methods and techniques – the “toolkit approach” (BELFIORE/BENNETT 2010: 121) – thus presuming that similar practices can be applied to different contexts. The decontextualized approaches inherent to the longstanding positivist paradigm tend also to minimize the role of human agency – both in artistic production and consumption – and neglect the centrality of ethics in evaluation processes (CHIARAVALLOTI/PIBER 2011).

At the same time, it is evident that a small number of works have been engaged in a quest for alternative approaches, probably as a counteraction to the dominance by positivism. The quest is for paradigms with a stronger emphasis on the intrinsic aspects of the individual experience, which are able to capture the contextual complexity of artistic and cultural activities. The case of the arts impact assessment, as a subgroup of the chosen works, illustrates some of these alternative approaches well.

On the one side, there are theory-based evaluations (GALLOWAY 2009), which draw on the observation that research on arts evaluation has focused on technical rather than on epistemological issues. It is argued that the generative understanding of causation underpinning theory-based evaluations is ontologically (better) suited to exploring the potential impact of the arts, as it can grasp unarticulated assumptions and values. On the other side, Belfiore and Bennet (2010) have proposed a humanities-based approach, building their argument on the observation that discourses around the transformative power of arts are based upon unquestioned assumptions, mostly aimed at legitimizing public spending. Their approach is located within a broad bundle of disciplines that explore the human condition and the products of human existence. As such, it can address questions of values and beliefs and capture the real effects – positive as well as negative – of artistic and cultural activities.

Pride and Prejudice. Coding-theme #9 captures the discussions relating to issues of implementation and elaborates on how stakeholders in ACOs reacted to the introduction of performance measurement or evaluation processes. This coding theme yields remarkable observations across region/language areas (coding-theme #1).

The English-written literature tends to reflect on the experiences during and after the implementation by emphasizing the benefits, the organizational learnings as well as the chances for future endeavors. This observation is conceptualized as ‘pride’. In the literature on German-speaking countries, these experiences are only occasionally discussed. Instead, another recurrent theme is observed – namely, an (assumed) a priori negative attitude of the involved parties in ACOs towards the issues of performance measurement and evaluation. This is illustrated with negatively connoted words like ‘Skepsis’, ‘Abwehr’ and ‘repressive Kontrolle’, which refer to skepticism, reluctance, defense, resistance, and even repressive control. The way in which studies in the German-speaking area describe, anticipate or expect reactions towards evaluative practices in ACOs is conceptualized as ‘prejudice’.

The 'prejudicial' observation is complex because it involves numerous voices: the ACOs, the researchers and authors of the reviewed works as well as the meta-ethnographer who is the author of this study (DOYLE 2003). Nonetheless, it raises many further relevant inquiries. First, it is unclear whether this perceived prejudicial attitude is based on anecdotal evidence and whether it indeed reflects the current discourse in arts management praxis or rather relies on outdated perceptions of it. Second, assuming this prejudicial attitude reflects the current arts management praxis, another line of inquiry relates to issues of funding, legitimacy, and control, and whether this difference in attitude reflects the different funding structures and, consequently, behavior and expectations across regions.

Good Cop, Bad Cop. Following coding theme #4, #7 and #8, it can be observed that while most of the studies are focused on the introduction of new (or adapted) evaluation techniques or performance measurement systems, only a few studies critically reflect on the existing literature and/or address the range of theoretical, epistemological, aesthetic, political, and ethical challenges embedded in evaluative practices. Coding-theme #7 intends to capture how, why, and by whom research endeavors were initiated and/or owned. Out of all the works analyzed, seldom did ACOs themselves proactively take the initiative to develop a performance measurement system or an evaluation model. Most of the projects seem to have been initiated from the researchers themselves, sometimes with funding from public or private bodies. However, it was not clear in all cases whether research was 'just' funded or whether in fact it was commissioned. A further remarkable observation is that it remains unclear how tangible results and evidence-based practice inform the formulation of cultural policy. Nevertheless, assuming some of the studies were indeed commissioned, it is still surprising that the literature focused mostly on instrumental and methodological issues, by introducing new approaches for performance measurement and evaluation, while other critical issues inherent to these practices were little investigated.

The discussion about approaches to performance measurement and evaluation becomes unavoidably entangled in considerations about funding and politics. This raises the question of the interplay between arts management research and cultural policy. In view of this, it seems that the arts management research agenda has attached itself to the cultural policy agenda, by following the so called performance measurement imperative (GSTRAUNTHALER/PIBER 2012) and fulfilling

the expectations of measurement (PATON 2003) that characterizes the realm of public policy. In other words, it could be suggested that arts management research has followed advocacy considerations by producing new and optimized approaches to performance measurement and evaluation that facilitate the legitimization of public spending for the arts and cultural sector. This suggestion yields the question: Why has arts management literature adopted a ‘good cop’ attitude rather than follow a more critical – bad cop-like – research agenda? This question takes the discussion to a higher level, as it enquires into the role of arts management research and its interplay with cultural policy.

4. Discussion

The following section will first discuss the aforementioned four findings and then offer an encompassing conclusion, reflecting on them from a theoretical perspective.

‘Third time’s a charm?’ manifests the emergence of an (overlapping) third development phase in performance measurement and evaluation of the arts and culture. The questions raised are linked to a growing criticism of the decontextualized approaches found within the dominant positivist paradigm, which was captured by *‘A paradigm on the move?’*. As the existing underlying assumptions are contested, it is natural that a string of new alternative approaches emerge, so this second title should also illustrate the quest for alternative paradigms that can better capture the context-related uniqueness of artistic and cultural activities and the broader environment in which they are embedded. In this light, the third phase is not only characterized by challenging the underlying foundations of multidimensional frameworks and contesting the dominant positivist paradigm, but also by a shift towards alternative paradigms.

As a side note, it is worth noting that in the context of the 9th Annual Conference of the *Fachverband Kulturmanagement*⁹ in January 2016, on the subject of evaluation in the arts and culture, alternative paradigms and new approaches were presented. This was the case, for

- 9 The *Fachverband Kulturmanagement/Association of Cultural Management* is a professional association of academic lecturers and researchers in the field of cultural management in German-speaking countries. Its mission is the representation and advancement of cultural management in teaching and research. For details, visit <<http://www.fachverband-kulturmanagement.org>>.

example, with the exploration of art-based methods (MAGKOU 2016) or a value and framework oriented on Schulz von Thun's four-sides model (SVENSSON 2016).¹⁰ These approaches are not included in this meta-synthesis because at the time of researching/writing they had not been published. However, they are brought here into the discussion since they support the observation concerning the emergence of the aforementioned third phase in the debate surrounding the performance measurement and evaluation of ACOs.

Pride and Prejudice illustrates the observation that whereas the international literature tends to emphasize the benefits and learnings involved during and after the implementation of evaluative practices, a rather prejudicial attitude in the discourse of German-speaking countries can be observed. This points to a perception mismatch between theory and practice. In other words, and referring to the aforementioned three developing phases, it can be suggested that the way ACOs perceive (or are perceived to perceive) the current academic debate on performance management and evaluation is somehow between the first and second phase.

The last title *Good Cop, Bad Cop* takes the performance measurement and evaluation debate to a higher level by questioning arts management research and its interplay with cultural policy. Although addressing these issues is well beyond the scope of this study, some interesting questions emerge, which are formulated below in a rather provocative manner in order to bring them up for discussion. Could the 'research agenda attaching' be understood as a symptom that manifests the 'need' for arts management research to actually legitimize itself (by supporting cultural policy)? Or, on the contrary, does the rather 'uncritical research agenda' in arts management actually suggest that legitimization is (still) needed?

4.1 Conclusion

To conclude this meta-synthesis, the findings will be connected to existing knowledge in a way that serves as an intellectual transition from a description or analysis of phenomena to a generalization (SWANSON 2013) by analyzing the above-mentioned four titles through the lens of system theoretical discourse. When a theoretical framework is appropriate to the findings, its application can result in great explanatory power, as it allows (novel) understandings to be used in more informed

¹⁰ These approaches are not included in this meta-synthesis because, at the time of researching/writing, they had not been published.

and effective ways (TRÖNDLE 2010: 35). System theoretical discourse has been chosen to reflect on the findings of this study because system theory provides the field of arts management with a theoretical foundation with which issues such as cultural policy, arts and cultural organizations, and the role of art in society can be analyzed. Additionally, a system theoretical perspective allows for a theoretical linkage between arts management and related research disciplines such as management and aesthetics (TRÖNDLE 2010).

According to Luhmann's social systems theory, modern society is composed of functional systems including the political, economic, scientific, religious, and art systems (LUHMANN 1984). Each system has its own self-valorizing logic, expressed through system-specific binary codes and programs, and is operationally closed and autonomous. The core element of operation is communication. For example, the economy is a self-perpetuating system of payments; the science system produces scientific communication in terms of true/false; and the political system generates collectively binding decisions. Just like any other functional system, the art system participates in society with operative closure, in this case by differentiating itself as a system (LUHMANN 2000: 134) that provides society with a distinct observation code based on aesthetic perceptions (TRÖNDLE 2010: 26).

According to Luhmann, a system will principally only understand and use its own code, and will not consider the code of another system. For example, science's version of what is true does not guarantee that this truth will be recognized by religion or politics. External demands to any given functional system that are expressed in other systems' codes will either be treated as irrelevant and thus ignored or be handled as an irritation to overcome. Each system models constraints differently, reflecting the relevance of each constraint to the system's own reproduction and conservation; this is known as system autopoiesis. Autopoiesis can be understood as self-preservation through self-reproduction of system-specific communication (TRÖNDLE 2010:29).

In view of the above, the 'prejudicial' reaction of ACOs in German-speaking countries to evaluative practices (*Pride and Prejudice*) might be considered a 'logical' reaction of the system. The reaction is intended to protect the system's operational autonomy to other systems by preserving its own mode of legitimation. The perception of evaluative practices as an irritating external demand to the system might be accentuated by the observation that performance measurement and evaluation tend to be dominated by decontextualized approaches inherent to the positivist

research tradition. As such, they are less able to capture the context-related uniqueness of artistic and cultural activities and the broader environment in which they are embedded.

Nonetheless, any given system is constrained by its external environment and its material dependence on the performance of other systems (relativization of autopoiesis). Luhmann's (1997) concept of structural coupling opens up the possibility of interdependence between systems as they communicate and (co-)evolve through interaction. These concepts may explain why, in the international literature – as opposed to the reactions found in German-language literature – responses to performance measurement and evaluation practices expressed in other systems' operating codes are not perceived as a threat to the system's operational autonomy, but rather as a necessary interaction with other systems. This is because ACOs in regions other than German-speaking areas, and in particular in Anglo-Saxon countries, are not primarily funded by public subsidies but mostly rely on private donations. This suggests an interdependence with others systems, particularly the economic system. This might explain why ACOs in Anglo-Saxon countries do not exhibit a 'prejudicial' reaction to evaluation practices, but rather 'proudly' interact with these external demands. As a final point, it could be mentioned that, depending on the degree of constraint or interdependence upon other systems, systems will ignore, react, or interact with external demands. Following this premise, both 'pride' and 'prejudice' are logical strategies for the system's self-perpetuation.

The aforementioned third development phase (*Third time's a charm?*) as well as the search for alternative paradigms to the positivist research tradition (*A paradigm on the move?*) can be understood as a quest towards performance measurement and evaluation approaches derived from the system's own specific operations and thus expressed by means of the art system's own codes. Such alternative approaches underlie the art system's self-valorizing logic and therefore may be perceived not as a threat but rather as a part of the self-perpetuating program. However, according to the above reflection on *Pride and Prejudice*, this quest for alternative paradigms seems to be more relevant in relation to contexts and environments in which the art system (and its subsystems) is less dependent upon and shows less structural coupling with other systems.

Lastly, the concept of relativization of autopoiesis may also be helpful in shedding some light on the interplay between arts management research and cultural policy (*Good Cop, Bad Cop*). Systems may not only

be interdependent and related through structural coupling, they may even be more deeply related through what is conceptualized as interpenetration. This occurs when a system presupposes the complex achievements of another system and can treat them as parts of its own operations. This is not to suggest that arts management research handles cultural policy operations as parts of its own system; indeed, this seems like too bold a proposition to make. However, it can indeed be suggested that both (sub-)systems are clearly related through structural coupling in terms of issues of agenda setting, legitimation, and funding, at least in the discourse around performance measurement and evaluation.

4.2 Contribution to theory and practice

The present study produces new insights and understandings that can shed light on future investigations and policy formulation in the issues of performance measurement and evaluation of arts and cultural activities as well as provide guidance to practitioners. The latter may be more interested in the study's practical implications for their own activities. Although the level of discussion of the study is quite theoretical, some managerial implications can nevertheless be drawn.

Following the premise that, depending on the constraints imposed by their external environments and the material dependence on other systems, any given system will ignore, react, or interact with external demands, it seems crucial for practitioners to choose the 'right' approach to performance measurement and evaluation. While context related, the 'right' approach should always aim at 'interaction'. This can be achieved either because performance measurement and evaluation is approached in terms of the system's own valorizing logic or because the (sub-)system being assessed, for example a particular ACO, is able to operate in terms of another system's code. If the chosen approach cannot achieve interaction, the ACO (and its subsystems) will either react to or ignore external demands. That is, performance measurement and evaluation requests will be undermined, sabotaged, or conducted as a pro forma exercise that satisfies external constituencies but adds little (or even negative) value to the internal practices of the ACO, particularly in the artistic realm. In this light, it may be sensible for public and private funding agencies to consider relying more on professional judgments based on experience and expertise than on performance measurement and evaluation practices that are less likely to achieve 'interaction'.

The study raises questions for researchers that call for future investigation as well as contributes to the understanding of the potential of me-

ta-syntheses within arts management research. In relation to the former, there is much research potential for studies that consider the artistic context and the broader cultural and social environment in which they operate. In particular, this potential lies in the exploration of alternative paradigms that develop evaluation practices from the system's own specific operations; that is, performance measurement and evaluation models approached from a system's self-valorizing logic and expressed in terms of the art system's codes. Such alternative models may be perceived not as an irritating external demand but rather as an element of the self-perpetuating program. There is also potential for empirical and interdisciplinary research in this area. For example, involving emergent art-based research approaches – or even artists – could be very fruitful for the advancement of these issues. Furthermore, considering the normative nature of the existing literature, approaches that consider the role of human agency in the production and consumption of art may help minimize the existing theory-praxis gap.

At the methodological level, the study contributes to the existing body of knowledge as it represents the first attempt to analyze the literature on performance measurement and evaluation in the arts and cultural nonprofit sector by conducting a meta-synthesis. The contribution resides in, among other things, the fact that this methodology offers the possibility of synthesizing and relating existing knowledge across time, research and artistic disciplines, methodological approaches, and space. And this is presented in a way that is not primarily additive or 'typological' (as, for example, a traditional literature review could achieve), but reinterpetative. The validity of a meta-synthesis does not depend upon replication logic but rather upon critical interpretation (DOYLE 2003).

Meta-syntheses can offer a historical staging, such as the analysis of the development phases identified in the performance measurement and evaluation literature (*Third time's a charm?*). They extend the borders of each reviewed study (DOYLE 2003), thus allowing an unravelling and bringing to the fore of new (aspects of) relationships, such as the interplay between arts management research and cultural policy (*Good Cop, Bad Cop*), and how this shapes the discourse surrounding the performance measurement and evaluation of arts and cultural activities. Meta-syntheses can also offer an explanatory context. For example, the final analysis of the findings through the lens of system theoretical discourse illuminate on the 'pride' and 'prejudice' strategies of ACOs during and after the implementation of evaluative practices (*Pride and Prejudice*).

Finally, meta-syntheses are currently proliferating in the international academic literature, particularly in research fields characterized by a strong focus on qualitative research, evidence-based practice, and evidence-based policy making (HASSELER 2007). For researchers wishing to synthesize studies in a systematic and rigorous way, this paper has suggested that this methodology is a promising one. Such syntheses can help to go beyond individual cases and add value to the body of knowledge in arts management research and the advancement of this discipline by not only “acknowledging the importance of the uniqueness of individual cases, but also the uniqueness of collectives” (DOYLE 2003: 340).

4.3 Limitations

In regard to the application of the chosen methodology for this study, some limitations arose initially within the context of the bibliographic search. This was induced by the attempt to analyze the literature across language regions, and in particular with respect to the translation of terms sheds light, as the same (translated) word may have different connotations in different languages. Secondly, previous meta-ethnographies, particularly within the Cochrane policy (HANNES/LOCKWOOD 2012), have recommended the use of at least two researchers to identify and assess literature, extract and cross-check material, and agree on coding schemes. The present study was conducted by one researcher; hence there was little opportunity to solve inconsistencies through peer discussion. Lastly, even though a great deal of importance was given to ensuring quality sampling – for example, by excluding textbooks and master theses from the analysis – the sample nevertheless encompasses peer-reviewed articles and (academic) books, such as dissertations.

Overall, the paper sets out to address a broad problem statement, the whole breadth of which the study could not possibly cover. The study has an exploratory character and was conducted as a meta-synthesis framed alongside Noblit and Hare’s (1988) seven-steps for meta-ethnographic inquiry (see the ‘Methodology’ section). As a starting point, this iterative approach takes a broad overarching research question framed by the problem statement. However, in this type of interactional interpretive process, as Denzin (1989) describes it, the formulation of the research question is not a simple sequential procedure but rather an iterative one that is likely to be sharpened over the course of the study. The advantage of this approach is that such research design allows for deeper questions to emerge as the inquiry evolves, leading to novel understandings of the reviewed literature. On the other hand, as an (almost inevitable) draw-

back, findings made in this way may not thoroughly address the initial problem statement.

Nonetheless, the original application of this methodology to arts management research, as well as the study's findings and their framing within the system theoretical discourse, can be considered as contributing novel insights and understandings at a theoretical, methodological, and practical level to the discourse of performance measurement and the evaluation of artistic and cultural activities at the noncommercial edge of the arts and cultural sector.

Contributor

Leticia Labaronne is a Senior Research Associate at the Center of Arts Management at ZHAW Zurich University of Applied Sciences, School of Management and Law. She studied Politics, Public and Nonprofit Management and has worked in the performing arts for over ten years. Her current doctoral research at Zeppelin University relates to evaluation practices in the performing arts.

References

- BADIA, Francesco/DONATO, Fabio (2013): Performance Measurement at World Heritage Sites: Per Aspera ad Astra. – In: *International Journal of Arts Management* 16/1, 20–33.
- BAECKER, Dirk (2011). *Einführung in die Systemtheorie*. Heidelberg: Auer.
- BECKER, Howard S. (1982): *Art Worlds*. Berkeley: California UP.
- BELFIORE, Eleonora/BENNETT, Oliver (2007): Rethinking the social impacts of the arts. – In: *International Journal of Cultural Policy* 13/2, 135–151.
- BELFIORE, Eleonora/BENNETT, Oliver (2010): Beyond the 'Toolkit Approach': Arts Impact Evaluation Research and the Realities of Cultural Policy Making. – In: *Journal for Cultural Research* 14/2, 121–142.
- BEHN, Robert D. (2001): *Rethinking Democratic Accountability*. Washington/DC: Brookings Institution.
- BETZLER, Diana (2015): Theatre Quality Frame – A System of Business Processes and Performance Monitoring for Theatres. – In: *Proceedings XIII International Conference on Arts and Cultural Management AIMAC 2015*, Aix-Marseille University.
- BROOKS, Arthur C. (2000). The 'Income Gap' and the Health of Arts Nonprofits. – In: *Nonprofit Management and Leadership* 10/3, 271–286.
- BROOKS, Michael/MILNE, Chris/JOHANSSON, Klas (2002): Using Stakeholder Research in the Evaluation of Organisational Performance. – In: *Evaluation Journal of Australasia* 2/1, 20–26.
- MCDONNELL, Ian (2006): *Events Management*. Oxford: Elsevier.
- CAIRNS, Ben/HARRIS, Margaret/HUTCHISON, Romaine/TRICKER, Mike (2004): *Improving Performance? The Adoption and Implementation of Quality Systems in UK Nonprofits*. A paper prepared for presentation to the 33rd Annual ARNOVA Conference, Los Angeles/CA.

- CHIARAVALLLOTI, Francesco (2014): Performance Evaluation in the Arts and Cultural Sector: A Story of Accounting at Its Margins. – In: *The Journal of Arts Management, Law, and Society* 44/2, 61-89.
- CHIARAVALLLOTI, Francesco/PIBER, Martin (2011): Ethical Implications of Methodological Settings in Arts Management Research: The Case of Performance Evaluation. – In: *The Journal of Arts Management, Law, and Society* 41/4, 240-266.
- CLEMENTS, Paul (2007): The Evaluation of Community Arts Projects and the Problems with Social Impact Methodology. – In: *International Journal of Art, Design and Education* 26/3, 325-335.
- COHEN, Claire/PATEN, Maldwyn (2000): Making a Meal of Arts Evaluation: Can Social Audit Offer a More Balanced Approach? – In: *Managing Leisure* 5/3, 103-120.
- CORDES, Jens/MANSCHWETUS, Uwe/SCHIMKUS, Katja (2008): Qualitätsmanagement. – In Geyer, Hardy (Ed.), *Kulturmarketing*. München, Wien: Oldenbourg
- COWEN, Tyler (2008): Good and Plenty: The Creative Successes of American Arts Funding. – In: *Journal of Economic Literature* 46/1, 189-192.
- GETZ, Donald (2005): *Event Management and Event Tourism*. New York: Cognizant Communication.
- DANTO, Arthur C. (1964): The Artworld. – In: *The Journal of Philosophy* 61/19, 571-584.
- DENZIN, Norman K. (1989): *Interpretive Interactionism*. Thousand Oaks: Sage.
- DENZIN, Norman K./LINCOLN, Yvonna S. (Eds.) (2011): *The SAGE Handbook of Qualitative Research*. Thousand Oaks: Sage
- DICKIE, George (1984): *The Art Circle*. New York: Haven.
- DONATO, Fabio (2008): Managing IC by antennae: evidence from cultural organizations. – In: *Journal of Intellectual Capital* 9/3, 380-394.
- DOYLE, Lynn H. (2003): Synthesis Through Meta-Ethnography: Paradoxes, Enhancements, and Possibilities. – In: *Qualitative Research* 3/3, 321-344.
- EBRAHIM, Alnoor (2010): *The Many Faces of Nonprofit Accountability*. Working Paper Nr. 10-069 Harvard Business School. <<http://www.hbs.edu/faculty/Publication%20Files/10-069.pdf>> [02.03.2015].
- FINOCCHIARO, Massimo/RIZZO, Ilde (2009): Performance Measurement of Heritage Conservation Activity in Sicily. – In: *International Journal of Arts Management* 11/2, 29-41.
- GALLOWAY, Susan (2009): Theory-based Evaluation and the Social Impact of the Arts. – In: *Cultural Trends* 18/2, 25-148.
- GILHESPY, Ian (1999): Measuring the Performance of Cultural Organizations: A Model. – In: *International Journal of Arts Management* 2/1, 38-52.
- GILHESPY, Ian (2001): The Evaluation of Social Objectives in Cultural Organizations. – In: *International Journal of Arts Management* 4/1, 48-57.
- GRAY, Clive (2002): Local Government and the Arts. – In: *Local Government Studies* 28/1, 77-90.
- GSTRAUNTHALER, Thomas/PIBER, Martin (2012): The Performance of Museums and Other Cultural Institutions: Numbers or genuine judgements? – In: *International Studies of Management & Organization* 42/3, 29-42.

- HADIDA, Allègre (2015): Performance in the Creative Industries. – In: Jones, Candance/ Lorenzen, Mark/Sapsed, Jonathan (eds.), *The Oxford Handbook of Creative Industries*. Oxford: UP, 219-247.
- HANNES, Karin/LOCKWOOD, Craig (2012): *Synthesizing Qualitative Research*. West-Sussex: Blackwell.
- HANNES, Karin/MACAITIS, Kirsten (2012): A Move to More Systematic and Transparent Approaches in Qualitative Evidence Synthesis: Update on a Review of Published Papers. – In: *Qualitative Research* 12/4, 402-442.
- HASSELER, Martina (2007): Systematische Übersichtsarbeiten in qualitativer Gesundheits- und Pflegeforschung – eine erste Annäherung. – In: *Pflege & Gesellschaft* 12, 249-262.
- HILL, Alison/SPITTLEHOUSE, Claire (2001): What Is Critical Appraisal? – In: 'What is...?' *bulletins* 3/2 <http://meds.queensu.ca/medicine/obgyn.20150116/pdf/what_is/WhatIsCriticalAppraisal.pdf> [02.12.2016].
- HOLDEN, John (2004). *Capturing Cultural Value*. London: Demos.
- HUTTON, Will (1995): *The State We're in*. London: Cape.
- JOWELL, Tessa (2004): Government and the value of culture. – In: *Leisure Manager* 22/7, 8-10.
- KRUG, Kersti/WEINBERG, Chuck (2004): Mission, Money, and Merit: Strategic Decision Making by Nonprofit Managers. – In: *Nonprofit Management and Leadership* 7/2, 119-136.
- LAMPEL, Joseph/LANT, Theresa/SHAMSIE, Jamal (2000): Balancing Act: Learning from Organizing Practices in Cultural Industries. – In: *Organization Science* 11/3, 263-269.
- LEE, Richard P./HART, Ruth I./WATSON, Rose M/RAPLEY, Tim (2015): Qualitative Synthesis in Practice: Some Pragmatics of Meta-Ethnography. – In: *Qualitative Research* 15/3, 334-350.
- LINDQVIST, Katja (2012): Effects of Public Sector Reforms on the Management of Cultural Organizations in Europe. – In: *International Studies of Management & Organization* 42/2, 9-28
- LUHMANN, Niklas (1984): *Soziale Systeme*. Frankfurt/M.: Suhrkamp.
- LUHMANN, Niklas (1997): *Die Kunst der Gesellschaft*. Frankfurt/M.: Suhrkamp.
- LUHMANN, Niklas (1997): *Die Gesellschaft der Gesellschaft*. Frankfurt/M.: Suhrkamp.
- LUHMANN, Niklas (2000): *Art as a Social System*. California: Stanford UP.
- MAGKOU, Martina (2016): *Exploring the Potential of Arts-Based Methods for Cultural Evaluation*. Paper presented at the 9th Annual Conference of the Fachverband Kulturmanagement. Conference Documentation <<http://www.zhaw.ch/storage/sml/institute-zentren/zkm/upload/events/tagungsdokumentation-9-jahrestagung-winterthur.pdf>> [18.04.2016].
- NEW CULTURE PROMOTION ACT [Kulturförderungsgesetz, KFG] (2011): SR. 442.2 (Article 30, §2) <<http://www.admin.ch/opc/de/federal-gazette/2011/2971.pdf>> [16.12.2014].
- NIELSEN, Steffen/EJLER, Nicolaj (2008): Improving Performance? – In: *Evaluation* 14, 171-192.
- NOBLIT, George W./HARE, R. Dwight (1988): *Meta-Ethnography: Synthesizing Qualitative Studies*. Newbury Park/CA: Sage.

- PATON, Rob (2003): *Managing and Measuring Social Enterprises*. London: Sage.
- PATTON, Michael Q. (*2015): *Qualitative Research and Evaluation Methods*. London: Sage.
- PAULUS, Odile (2003): Measuring Museum Performance: A Study of Museums in France and the United States. – In: *International Journal of Arts Management* 6/1, 50-63.
- POPE, Catherine/MAY, Nicholas (*2006): *Qualitative Research in Health Care*. Oxford: Blackwell.
- SANDELOWSKI, Margarete/BARROSO, Juliet (2007): *Handbook for Synthesizing Qualitative Research*. New York: Springer.
- SCHEYTT, Oliver/ZIMMERMANN, M. (2006): Qualitätsmanagement in Kultureinrichtungen. – In: Loock, Friedrich/Scheytt, Oliver (eds.), *Kulturmanagement und Kulturpolitik: Die Kunst, Kultur zu ermöglichen*. Berlin: Raabe, 1-22.
- SCHOBER, Christian/RAUSCHER, Olivia/MILLNER, Reinhard (2012): Evaluation und Wirkungsmessung. – In: Simsa, Ruth/Meyer, Michael/Badelt, Christoph (Eds.), *Handbuch der Nonprofit-Organisation*. Stuttgart: Schäffer-Poeschel.
- SOREN, Barbara J. (2000): The Learning Cultural Organization of the Millennium: Performance Measures and Audience Response. – In: *International Journal of Arts Management* 2/2, 40-49.
- SVENSSON, Jenny (2016): *Die Kunst, Kultur (nicht nur) zu messen: Forschungsstudie zu Evaluationsmethoden und -Prozessen an Theatern der Region Skåne/Schweden*. Paper presented at the 9th Annual Conference of the Fachverband Kulturmanagement. Conference Documentation <<http://www.zhaw.ch/storage/sml/institute-zentren/zkm/upload/events/tagungsdokumentation-9-jahrestagung-winterthur.pdf>> [18.04.2016].
- SWANSON, Richard (2013): *Theory Building in Applied Disciplines*. San Francisco: Berrett-Koehler.
- TRÖNDLE, Martin (2010): Systemtheorie, ein Versuch [An attempt at system theorie]. – In: *Jahrbuch Kulturmanagement* 2010/1, 13-41.
- TSCHÄPE, Ruth/BRÜGGERHOFF, Stefan (2004): Qualitätsmanagement im Museum. – In: Heinze, Thomas (Ed.), *Neue Ansätze im Kulturmanagement*. Wiesbaden: Springer, 107-129.
- TURBIDE, Johanne/HOSKIN, Keith (1999): Managing Non-Profit Arts Organizations Through Management Accounting Systems: Mission Possible? – In: *International Journal of Arts Management* 1/2, 68-81.
- TURBIDE, Johanne/LAURIN, Claude (2009): Performance Measurement in the Arts Sector: The Case of the Performing Arts. – In: *International Journal of Arts Management* 11/2, 56-70.
- WEINSTEIN, Larry/BUKOVINSKY, David (2009): Use of the Balanced Scorecard and Performance Metrics to Achieve Operational and Strategic Alignment in Arts and Cultural Not-for-Profits. – In: *International Journal of Arts Management* 11/2, 42-55.
- WICKS, Andrew C./FREEMAN, Edward (1998): Organization Studies and the New Pragmatism: Positivism, Anti-Positivism, and the Search for Ethics. – In: *Organization Science* 9/2, 123-40.
- WIMMER, Ulla (2004): *Kultur messen: Zählen, Vergleichen und Bewerten im kulturellen Feld*. Berlin: Logos.
- WILLIAMS, Michael/BOWDIN, Glenn A.J. (2007): Festival Evaluation: An Exploration of Seven UK Arts Festivals. – In: *Managing Leisure* 12/2-3, 187-203.

- YOKER, Brandon W. (2010): The Logic of Evaluation and Not-for-Profit Arts Organizations: The perspective of an evaluation consultant. – In: *International Journal of Arts Management* 12/3, 4-12.
- ZORLONI, Alessia (2012): Designing a Strategic Framework to Assess Museum Activities. – In: *International Journal of Arts Management* 14/2, 31-47.
- ZULAUF, Jochen (2012): *Aktivierendes Kulturmanagement. Handbuch Organisationsentwicklung und Qualitätsmanagement für Kulturbetriebe*. Bielefeld: transcript.

Appendix

Year	Author	Title	Source
1986	Rosalie Hallbauer	<i>How Orchestras Measure Internal Performance</i>	Management Accounting
1986	François Matarasso	<i>Defining values: Evaluating arts programmes</i>	<i>The Social Impact of the Arts</i> 30 (Comedia)
1995	Thompson George	<i>Problems with service performance reporting: The case of public art galleries</i>	<i>Financial Accountability & Management</i> 11/4, 267-424
1997	J. Mark Schuster	<i>The Performance of Performance Indicators in the Arts</i>	<i>Nonprofit Management and Leadership</i> 7/3, 253-269
1999	Ian Gilhespy	<i>Measuring the Performance of Cultural Organizations: A Model</i>	<i>International Journal of Arts Management</i> 2/2, 38-52
1999	Johanne Turbide/ Keith Hoskin	<i>Managing Non-Profit Arts Organizations Through Management Accounting Systems: Mission Possible?</i>	<i>International Journal of Arts Management</i> 1/2, 68-81
2000	Barbara J. Soren	<i>The Learning Cultural Organization of the Millennium: Performance Measures and Audience Response</i>	<i>International Journal of Arts Management</i> 2/2, 40-44
2000	Claire Cohen/Maldwyn Pate	<i>Making a meal of arts evaluation: can social audit offer a more balanced approach?</i>	<i>Managing Leisure</i> 5/3, 103-120
2001	Ian Gilhespy	<i>The Evaluation of Social Objectives in Cultural Organisations</i>	<i>International Journal of Arts Management</i> 4/1, 48-57
2003	Odile Paulus	<i>Measuring Museum Performance: A Study of Museums in France and the United States</i>	<i>International Journal of Arts Management</i> 6/1, 50-63
2003	Victor Ginsburgh	<i>Awards, success, and aesthetic quality in the arts</i>	<i>Journal of Economics Perspectives</i> 17/2, 99-112
2004	Kersti Krug/ Charles B. Weinberg	<i>Mission, Money, and Merit: Strategic Decision Making by Nonprofit Managers</i>	<i>Nonprofit Management & Leadership</i> 14/3, 325-342
2004	Ruth Tschäpe/ Stefan Brüggerhoff	<i>Qualitätsmanagement im Museum [Quality Management in Museums]</i>	Heinze, T. (Ed.), <i>Neue Ansätze im Kulturmanagement [New approaches to cultural management]</i> , Wiesbaden: Springer, 107-129

Year	Author	Title	Source
2004	Sabine Boerner	<i>Artistic Quality in an Opera Company: Toward the Development of a Concept</i>	<i>Nonprofit Management and Leadership</i> 14/4, 425-436
2006	Oliver Scheytt/ Michael Zimmermann	<i>Qualitätsmanagement in Kultureinrichtungen</i> [Quality Management in arts and cultural organizations]	Loock, F. & Scheytt, O. (Eds.), <i>Kulturmanagement und Kulturpolitik: Die Kunst, Kultur zu ermöglichen</i> [Cultural management and cultural policy: the art of cultural support]. Berlin: Raabe, 1-22
2007	Michael Williams/ Bowdin Glenn A.J	<i>Festival evaluation: An exploration of seven UK arts festivals</i>	<i>Managing Leisure</i> 12/2-3, 187-203
2007	Thomas Gstraunthaler/ Martin Piber	<i>Performance Measurement and Accounting: Museum in Austria</i>	<i>Museum Management and Curatorship</i> 22/4, 361-375
2007	Eleonora Belfiore/ Oliver Bennett	<i>Rethinking the Social Impacts of the Arts.</i>	<i>International Journal of Cultural Policy</i> 13/2, 135-151
2007	Paul Clemens	<i>The Evaluation of Community Arts Projects and the Problems with Social Impact Methodology</i>	<i>International Journal of Art & Design Education</i> 26/3, 325-335
2008	Kate Clark/ Gareth Maeer	<i>The Cultural Value of Heritage: Evidence form the Heritage Lottery Fund</i>	<i>Cultural Trends</i> 17/1, 23-56
2008	Sabine Boerner/ Sabine Renz	<i>Performance Measurement in Opera Companies: Comparing the Subjective Quality Judgements of Experts and Non-experts</i>	<i>International Journal of Arts Management</i> 10/3, 21-37
2008	Sabine Boerner/ Hans Neuhoff Sabine Renz/ Volker Moser	<i>Evaluation in Music Theater: Empirical Results on Content and Structure of the Audience's Quality Judgment</i>	<i>Empirical Studies of the Arts</i> 26/1, 15-35
2008	Cordes, Jens et al.	<i>Qualitätsmanagement</i> [Quality management]	Geyer, H (Ed.), <i>Kulturmarketing</i> [Marketing in the arts and culture]. München, Wien: Oldenbourg

Year	Author	Title	Source
2008	Nora Wegner	<i>Vorab-Evaluation als Antwort auf aktuelle Herausforderungen an Museen</i> [Ex ante evaluations as key to address current challenges in museums]	<i>Keller, Rolf/Schaffner, Brigitte/Seeger, Bruno (Hrsg.): Spielplan. Schweizer Jahrbuch für Kulturmanagement 2007/2008</i> [Swiss Yearbook of Cultural Management] Bern: Haupt
2008	Susannah Eckersley	<i>Supporting Excellence in the Arts: From Measurement to Judgement</i>	<i>Cultural Trends</i> 17/3, 183-187
2009	Larry Weinstein/ David Bukovinsky	<i>Use of the Balanced Scorecard and Performance Metrics to Achieve Operational and Strategic Alignment in Arts and Culture Not-for-profits</i>	<i>Museum Management and Curatorship</i> 22/4, 361-375
2009	Massimo Finocchiaro/ Risso Ilde	<i>Performance Measurement of Heritage Conservation Activity in Sicily</i>	<i>International Journal of Arts Management</i> 11/2, 29-41
2009	Susan Galloway	<i>Theory-based Evaluation and the Social Impact of the Arts.</i>	<i>Cultural Trends</i> 18/2, 125-148
2009	Johanne Turbide/Claude Laurin	<i>Performance Measurement in the Arts Sector: The Case of the Performing Arts</i>	<i>International Journal of Arts Management</i> 11/2, 56-70
2010	Dagmar Abfalter	<i>Das Unmessbare messen?</i> [Measuring the unmeasurable?]	Wiesbaden: VS
2010	Miranda Boorsma/ Francesco Chiaravalloti	<i>Arts Marketing Performance: An Artistic-Mission-Led Approach to Evaluation.</i>	<i>The Journal of Arts Management, Law, and Society</i> 40/4, 297-317
2010	Brandon W Youker	<i>The Logic of Evaluation and Not-for-Profit Arts Organizations: The Perspective of an Evaluation Consultant</i>	<i>International Journal of Arts Management</i> 12/3, 4-12
2010	Eleonora Belfiore/ Oliver Bennett	<i>Beyond the "Toolkit Approach": Arts Impact Evaluation Research and the Realities of Cultural Policy Making</i>	<i>Journal for Cultural Research</i> 14/2, 121-142
2011	Gesa Birnkrant	<i>Evaluation im Kulturbetrieb</i> [Evaluating art organizations]	Wiesbaden: VS.
2011	Francesco Chiaravalloti/ Martin Piber	<i>Ethical Implications of Methodological Settings in Arts Management Research: The Case of Performance Evaluation</i>	<i>The Journal of Arts Management, Law, and Society</i> 41/4, 240-266

Year	Author	Title	Source
2012	Alessia Zorloni	<i>Designing a Strategic Framework to Assess Museum Activities</i>	<i>International Journal of Arts Management</i> 14/2, 31-72
2012	Patricia Munro	<i>Den ganzen Eisberg wahrnehmen. Die Förderung einer „partizipativen Kultur“ in Museen und welche Rolle Evaluation dabei spielen kann</i> [Taking into account the whole picture: The role of evaluations in advancing a participatory culture in museums]	Gesser, Susanne et al. (Hg.), <i>Das partizipative Museum</i> . Bielefeld: transcript, 213-220
2012	Jochen Zulauf	<i>Aktivierendes Kulturmanagement. Handbuch Organisationsentwicklung und Qualitätsmanagement für Kulturbetriebe</i> [Activating cultural management Handbook for organisational development and quality management in the arts and culture]	Bielefeld: transcript
2012	Thomas Gstraunthaler/Martin Piber	<i>The Performance of Museums and Other Cultural Institutions</i>	<i>International Studies of Management and Organization</i> 42/2, 29-42
2013	Sabine Hirschle	<i>Die Balanced Score Card</i>	Petra Schneidenwind (ed.), <i>Controlling im Kulturmanagement: Eine Einführung</i> [An introduction to controlling in arts management]. Wiesbaden: Springer VS, 126-142
2013	Vera Hennefeld/Reinhard Stockmann (Hrsg)	<i>Evaluation in Kultur und Kulturpolitik: eine Bestandaufnahme</i> [Evaluation in the arts and in cultural policy: Current state of knowledge]	Munster: Waxmann
2013	Francesco Badia/Fabio Donato	<i>Performance Measurement at World Heritage Sites: Per Aspera ad Astra</i>	<i>International Journal of Arts Management</i> 16/1, 20-33
2014	Francesco Chiaravalloti	<i>Performance Evaluation in the Arts and Cultural Sector: A Story of Accounting at Its Margins.</i>	<i>The Journal of Arts Management, Law, and Society</i> 44/2, 61-89

Year	Author	Title	Source
2014	Katya Johanson/Hillary Glow/Anne Kershaw	<i>New Modes of Arts participation and the Limits of Cultural Indicators for Local Government</i>	<i>Poetics</i> 43, 43-59
2015	Allègre Hadida	<i>Performance in the Creative Industries</i>	Jones, Candance/Lorenzen, Mark/Sapsed, Jonathan (eds.), <i>The Oxford Handbook of Creative Industries</i> . Oxford: UP, 219-247.

